

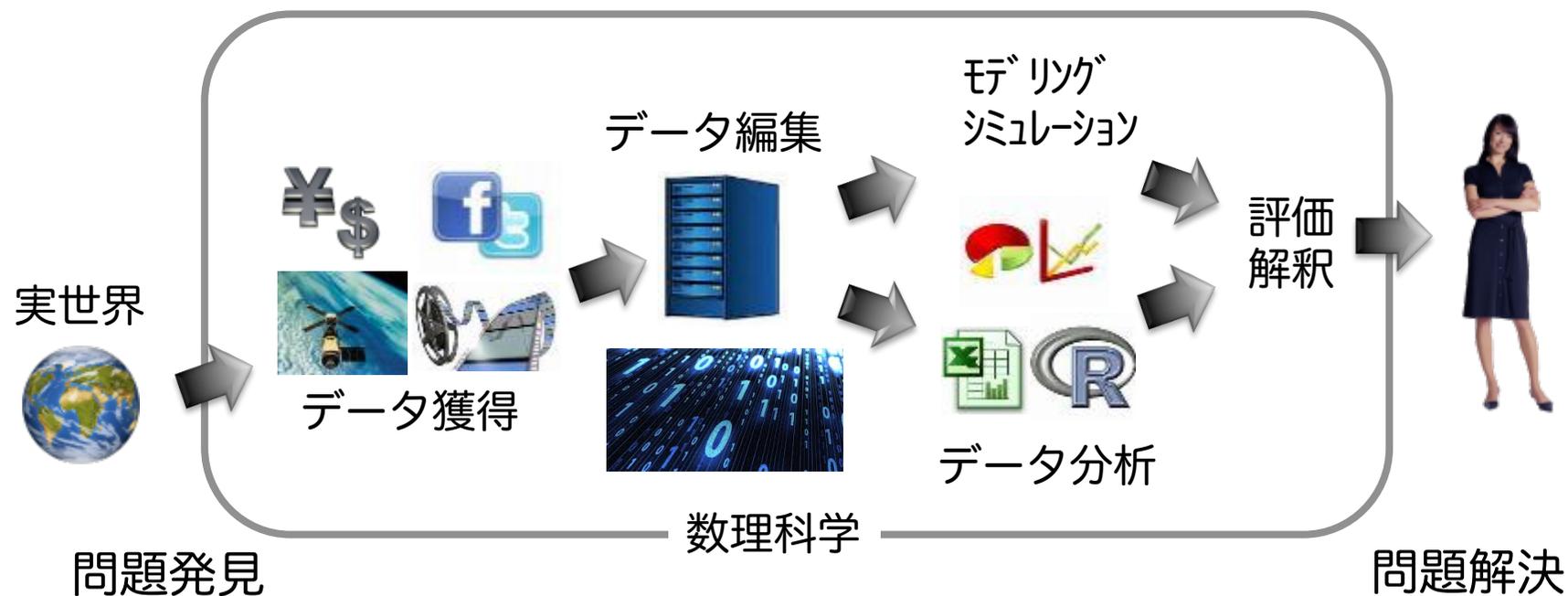
ベイズ統計

古谷知之

講義概要

- データ科学の動向
- なぜベイズ統計なのか？
- 授業スケジュール
- 成績判定方法

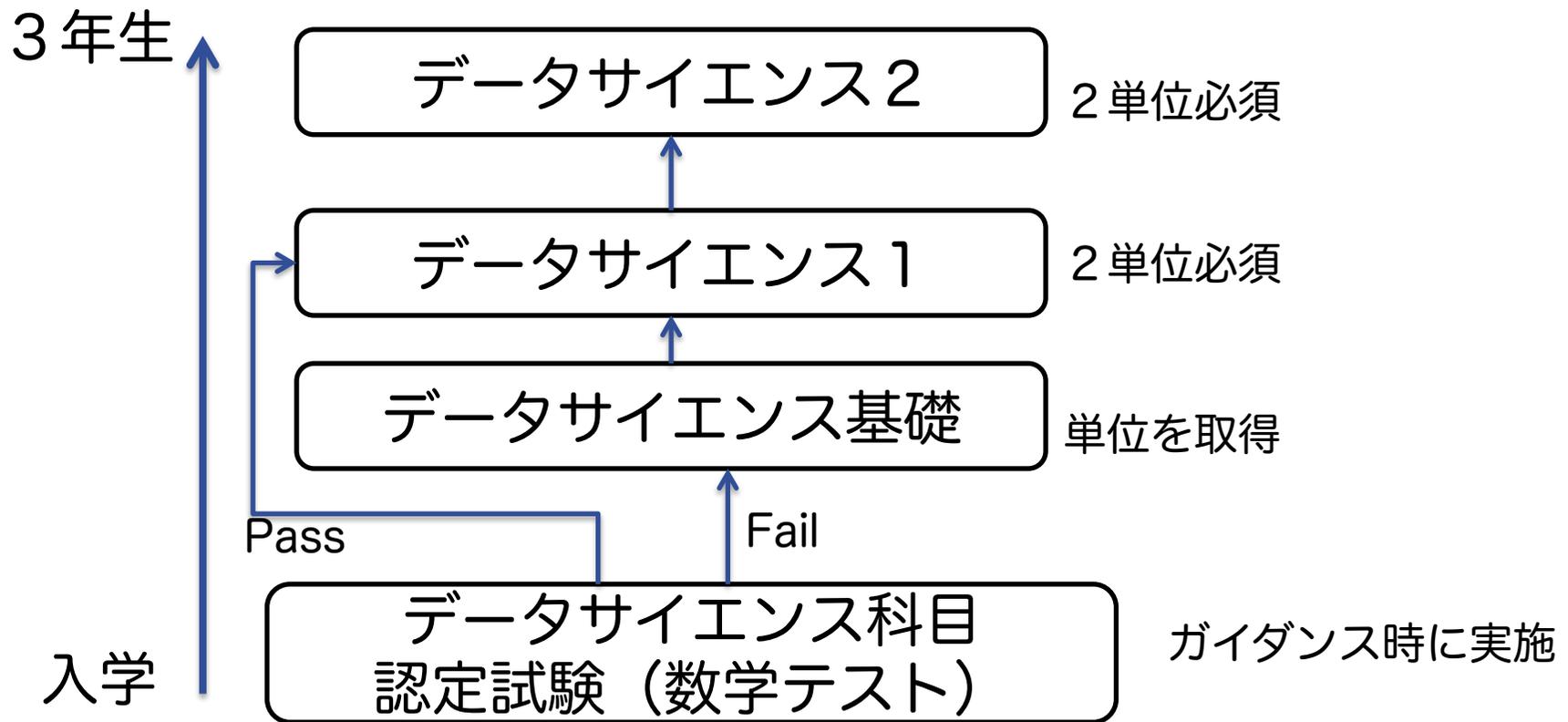
データサイエンスとは



データサイエンス科目とは

- いくつかもの分野に広く応用できる共通の基礎技法を習得するための科目であり、データの獲得・編集・分析手法とモデリング・シミュレーションおよび論理・確率・代数などの数理科学について学びます。
- データサイエンス科目では卒業後に最低限必要なスキルを提供しています。データサイエンス科目以外にも応用的なスキル科目が提供されているので、ニーズに合わせて履修してください。

データサイエンス科目の履修ステップ（新1年生）



データ/インフォメーション/インテリジエンス

インテリジエンス



インフォメーション



データ



実世界



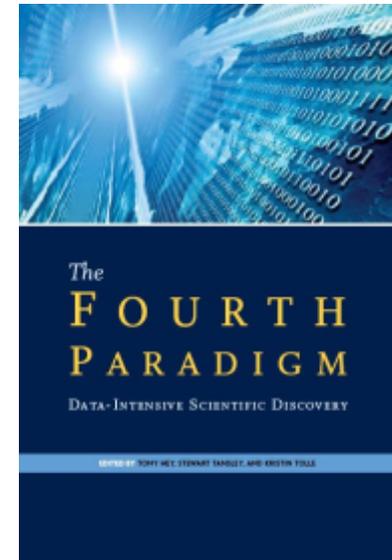
データ科学に期待される役割

- 政策の質向上
 - 予測・分析・モニタリング
精度改善
 - 課題解決の迅速化
- 政策サービス改善/効率化
 - 生活者・地域（患者）視点からの
政策
 - パーソナル化、オーダーメイド化
 - 情報公開、プライバシー保護



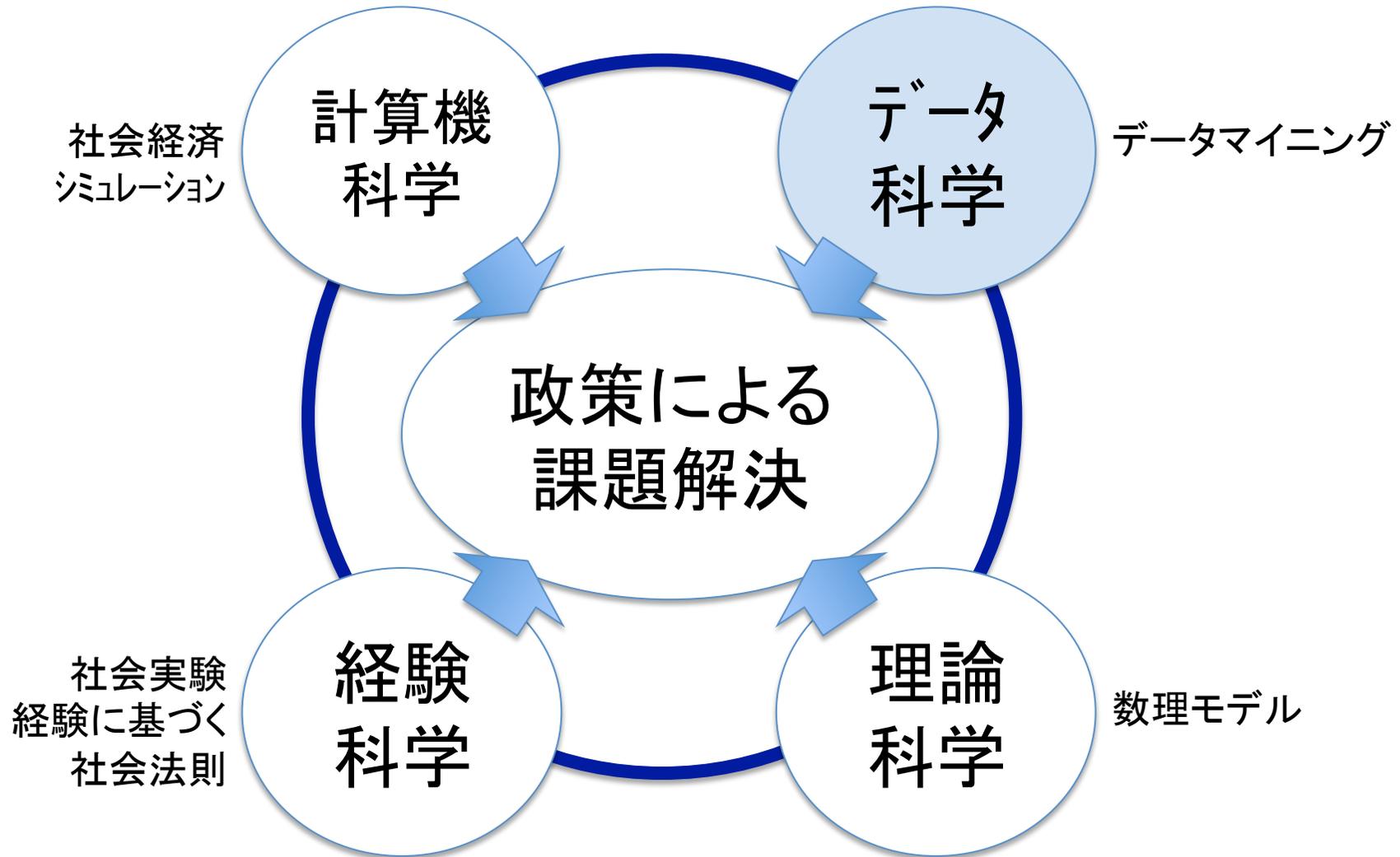
21世紀はデータサイエンスの時代

- 「ビッグデータ」や「オープンデータ」という言葉を一度は聞いたことがあると思います
- 今まで以上に多様で膨大なデータの解析が可能になり、データサイエンスは21世紀学問の主流を占めるといわれています
- データサイエンスに精通した人材育成が世界的な課題となっています

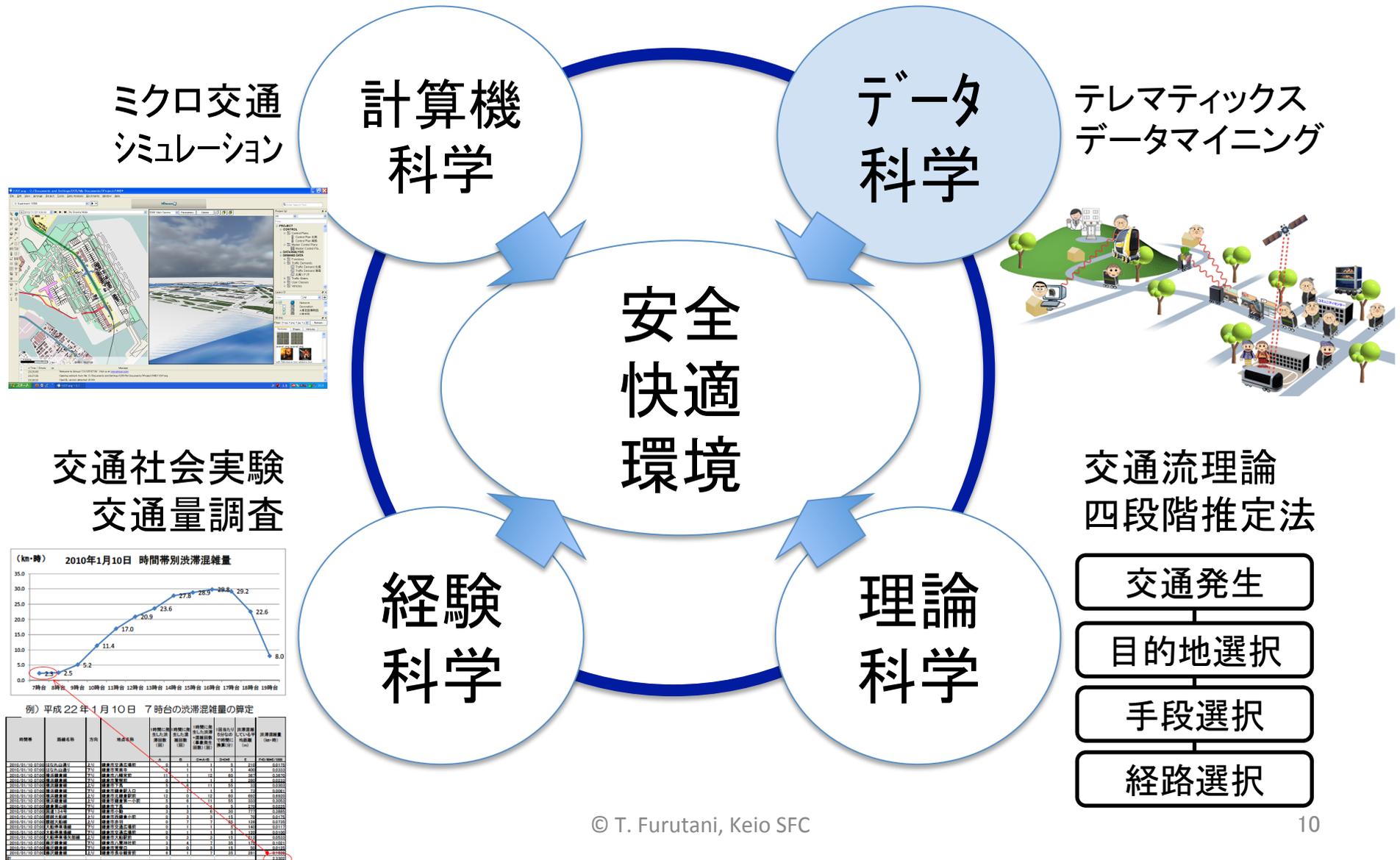


T. Hey, S. Tansley and K. Tolle (2009) "The Fourth Paradigm: Data-Intensive Scientific Discovery", Microsoft Research.

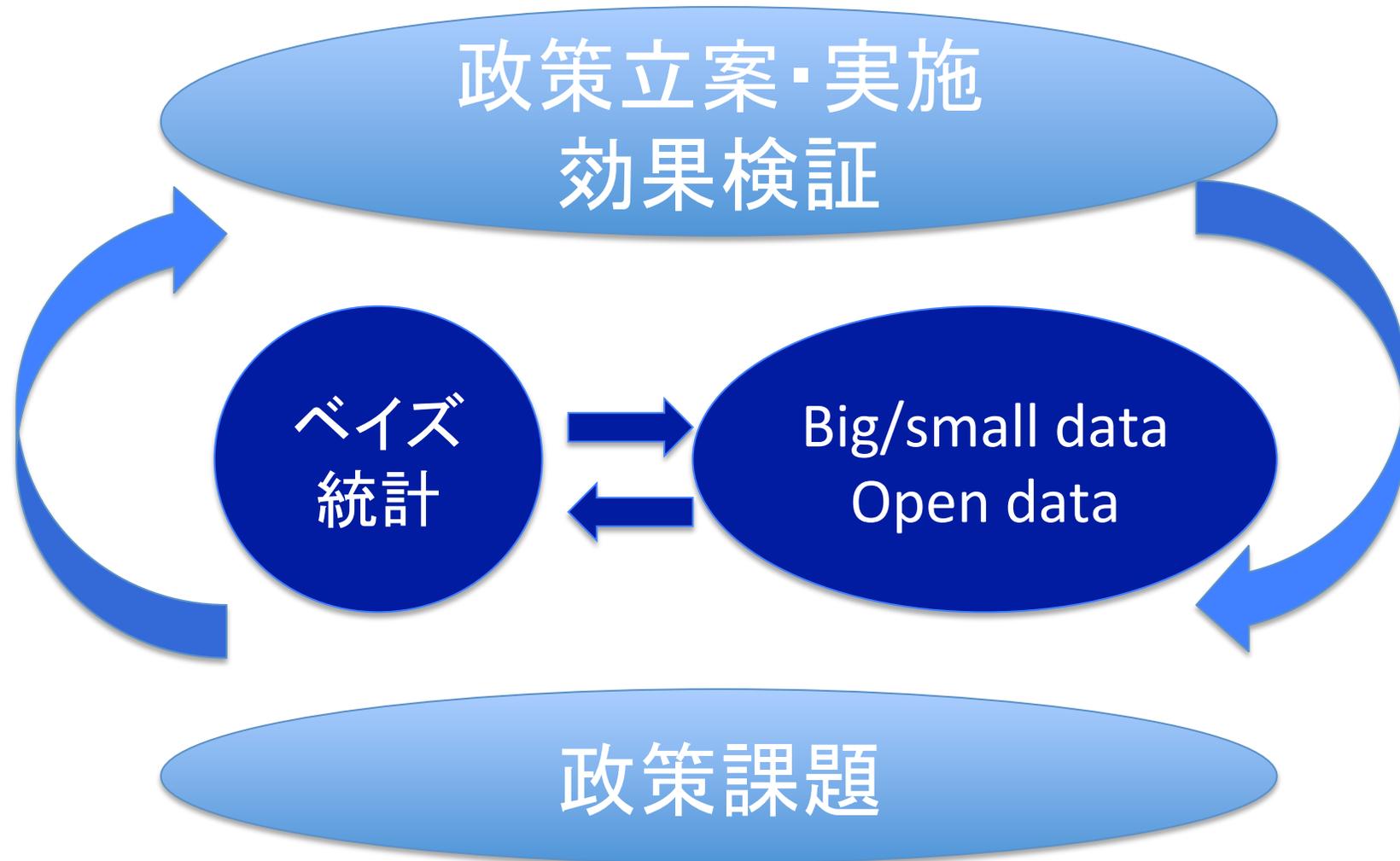
「4つの科学」と政策学



「4つの科学」と政策学：交通政策の例



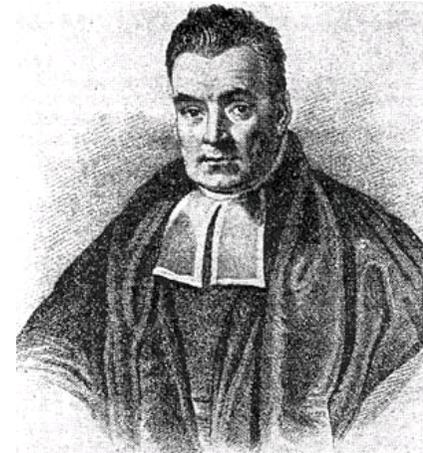
データ科学時代の政策立案



Thomas Bayes (1701?-1761)

<http://plus.maths.org/>

- 英国非国教会派の牧師
- 死後に、初期の条件付き確率に関する論文が発見され、遺作として発表される
 - 「事後」で条件づけられた「事前」の確率を計算できること = 「ベイズの定理」を発見



近現代統計学の系譜

画像は全てWikipediaから転載



ナポレオン

閣僚に
起用

追放

対立

ラプラス
微分方程式
予測確率



19c中

ガウス
正規分布
最小二乗法



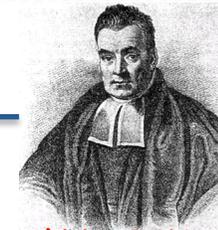
発見

ベイズ統計学 (初期)

天才による統計学

確率論重視!

ベイズ 18c
ベイズの定理
牧師



19c末~20c初



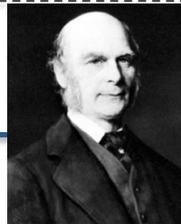
ナイチンゲール
医学統計学

貴族による統計学
(アカデミックリート)
大標本重視!

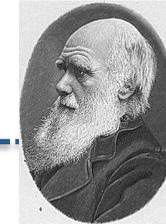
いとこ

結婚

ゴルトン
優生学
相関係数



いとこ

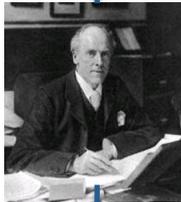


ダーウィン
生物学

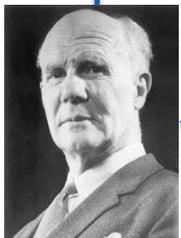
庶民による統計学 (非アカデミックリート)

数理統計学
記述統計学

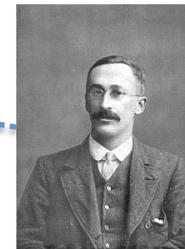
ピアソン
(父)
線形回帰
 χ^2 検定



ピアソン
(子)
仮説検定
信頼区間



対立



ゴセット
Student's T
ギネスの技師

小標本
重視!



フィッシャー
優生学
Z検定, p値
実験計画法
分散分析

20c後~21c

ベイズ統計学
(神再臨)

主観確率重視!



推測統計学

客観確率重視!

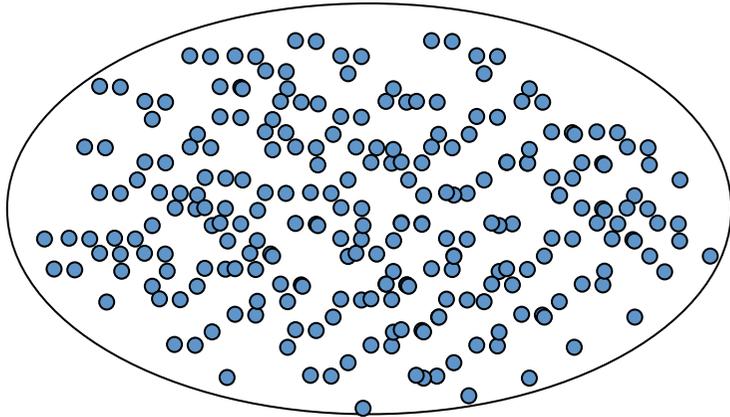
21世紀は確率論とベイズの時代

- 20世紀まで
 - 「データが十分に得られない時代」
 - データが少ないので数学的理論を精緻化
 - 古典統計、頻度主義統計、微積分
- 21世紀
 - 「大量のデータが得られる時代」
 - データの分布を考えた分析手法
 - 確率論、ベイズ統計



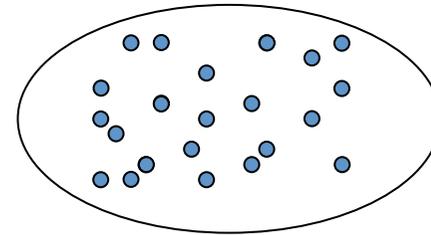
データサイエンスの基本的な考え方

データが大量に得られる場合



サンプルデータをそのまま
データ全体としてとらえる

データが大量に得られない場合



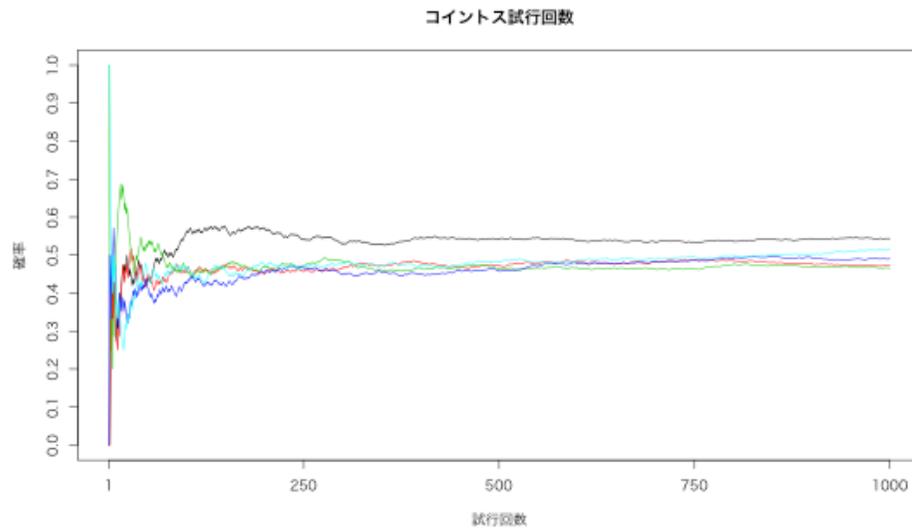
データ全体の形を推測
→推測統計学

サンプル数自体は特に気にしない
データ全体の形をシミュレーション
→ベイズ統計学

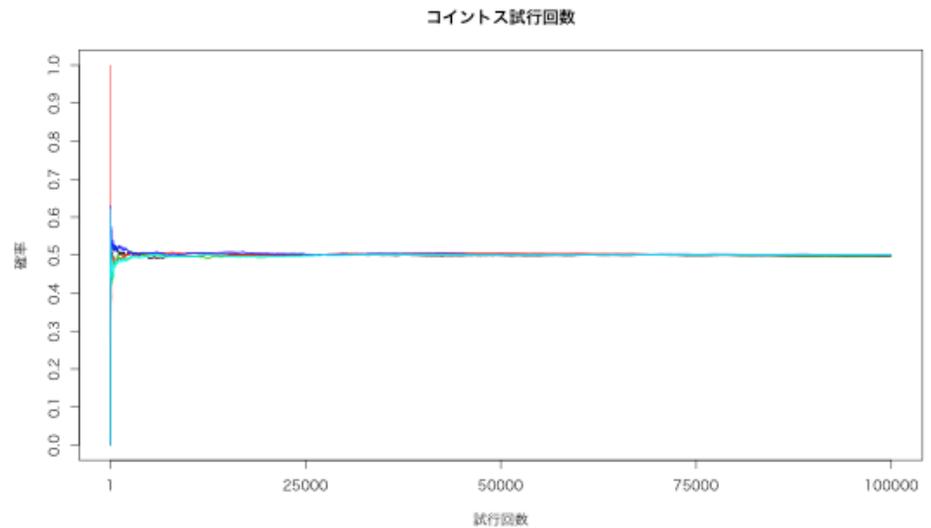
大量データを使えるようになると…

- コイントスをすると表が出る確率は1/2？

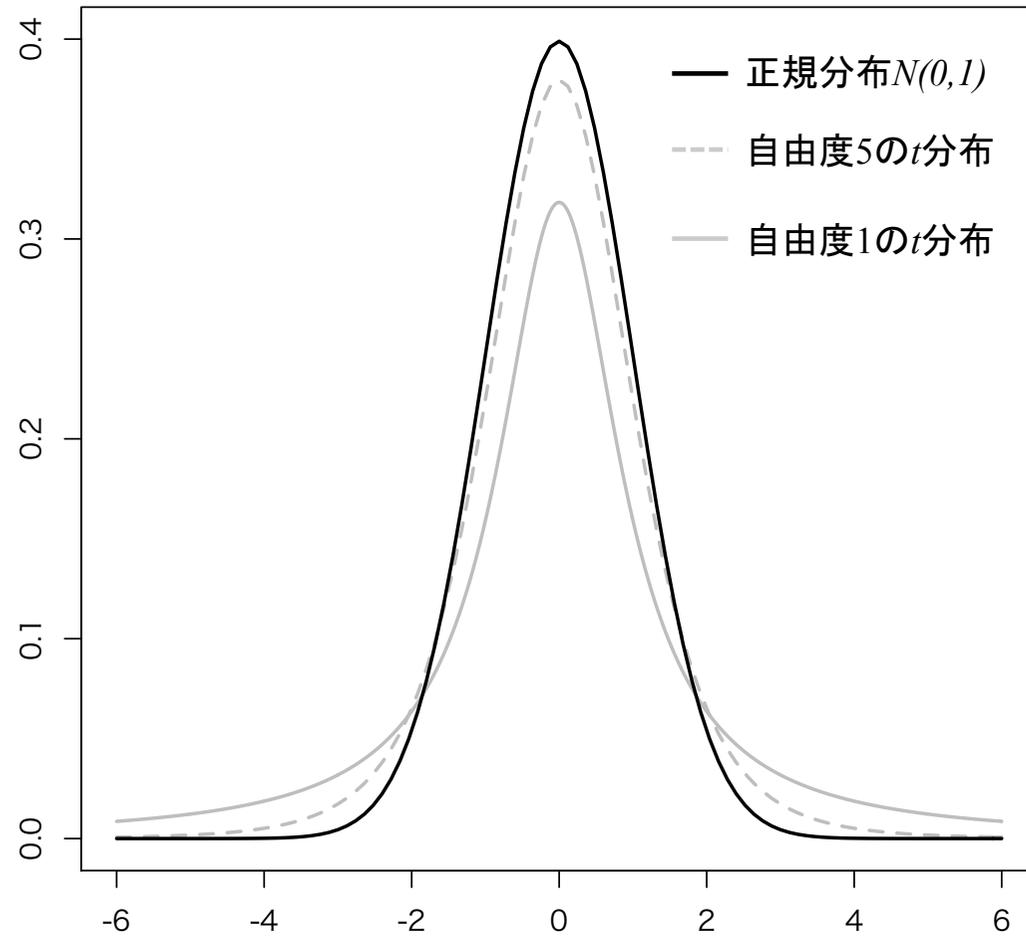
5人が1000回コイントス



5人が100000回コイントス

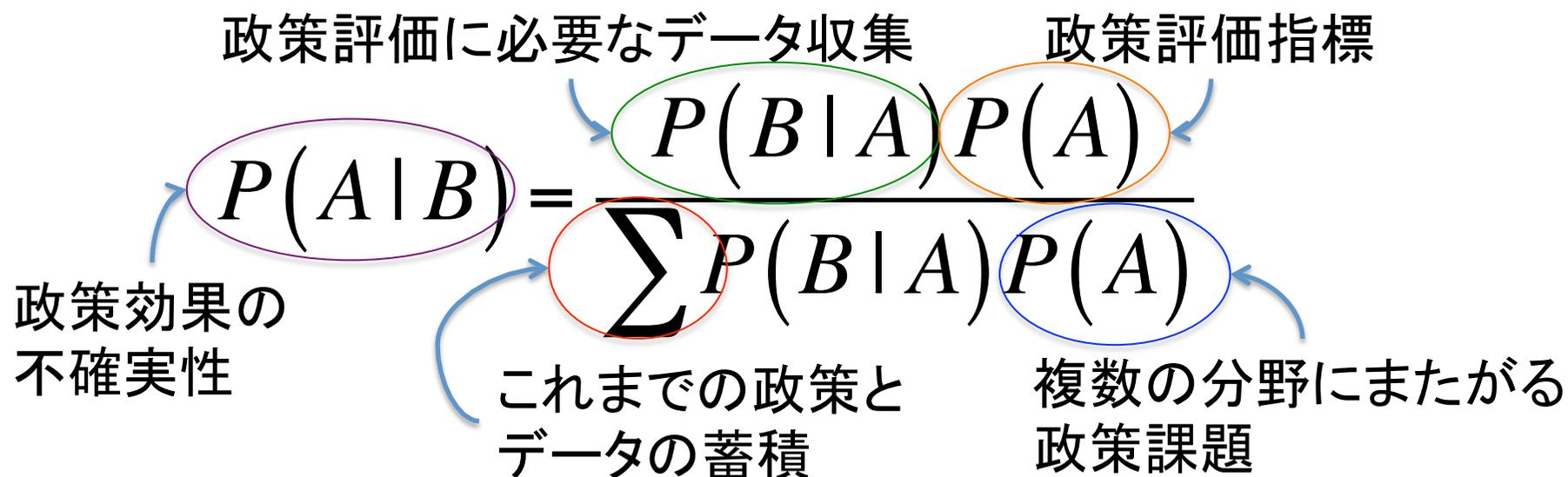


大量データを使えるようになると…



なぜ今、ベイズ統計なのか？

- 政策効果(A)を示す政策知見とデータ(B)
- 異なる分野のデータと政策知見を、比較的大量に組み合わせることが出来る
- 不確実性を数値化（確率化）できる



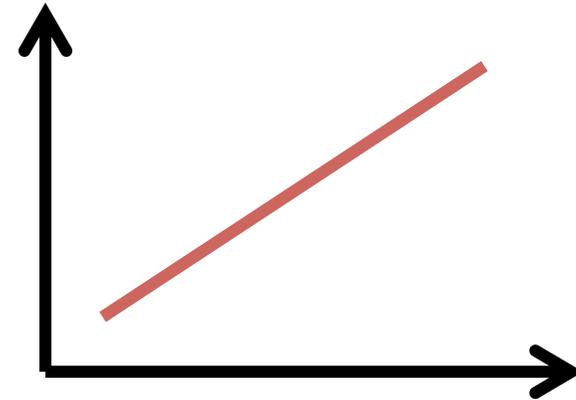
古典統計とベイズ統計の違い

古典統計：データをモデルにあわせる

ベイズ統計：データにあわせて自由にモデリング



データ



モデル

抽出



実世界

理論



ベイズ統計ができるとよい理由

Average Salary of Jobs with Titles Matching Your Search

[applied bayesian statistics](#) \$100,000

[applied statistics](#) \$59,000

[statistics](#) \$56,000

In USD as of Apr 4, 2014

40k

80k

120k

http://www.simplyhired.com/a/salary/search/q-applied%20bayesian%20statistics/q_1-applied+statistics/q_2-statistics

ベイズ統計が用いられる例

- 選挙の予測（政治学）
- 金融オペレーションリスク管理（経済学）
- 新薬の治験効果検証（医療・バイオ）
- 生物多様性の分析（生態学）
- 犯罪捜査
- スポーツ選手の行動履歴
- :
- ベイズ統計を使わないと論文が認められない分野も→ベイズさえ身につければanalystとしての職に困らない

シャーロック・ホームズはベイジアン？

NHK 海外ドラマ ホームページ

ご意見・ご感想

FAQ

DVDや書籍など

スタッフブログ

NHK番組表

NHK海外ドラマラインナップ

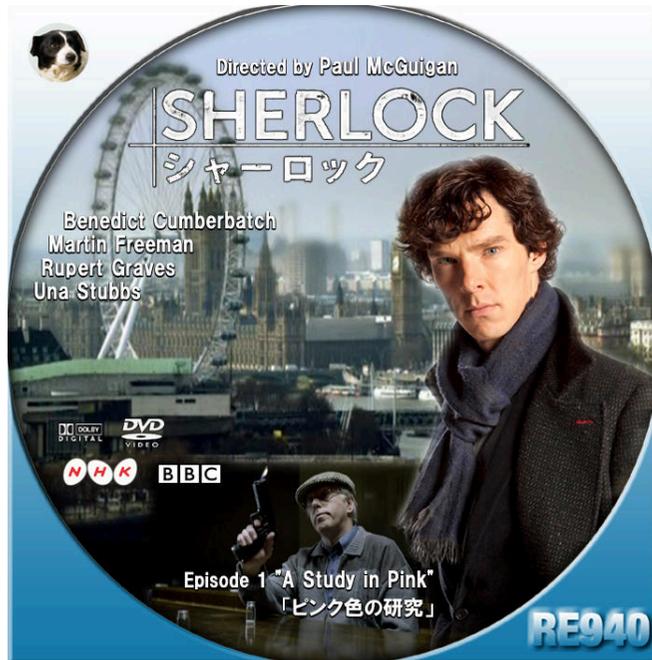
新シーズン
放送決定!
SHERLOCK 3
シャーロック
詳しくはこちらから▶

トップページ 第1シーズン 第2シーズン 登場人物 特別番組 インタビュー クイズ

特別番組のお知らせ

「徹底解明！SHERLOCKの秘密 <仮>」 BSプレミアム 5月18日(日) 午前0時30分～[*17日(土) 深夜0時30分～]

キャストインタビューも満載！



第1シーズン【全3回】BSプレミアム

2014年4月6日(日)、13日(日)、20日(日) 午前0時30分～[*土曜深夜]

第2シーズン【全3回】BSプレミアム

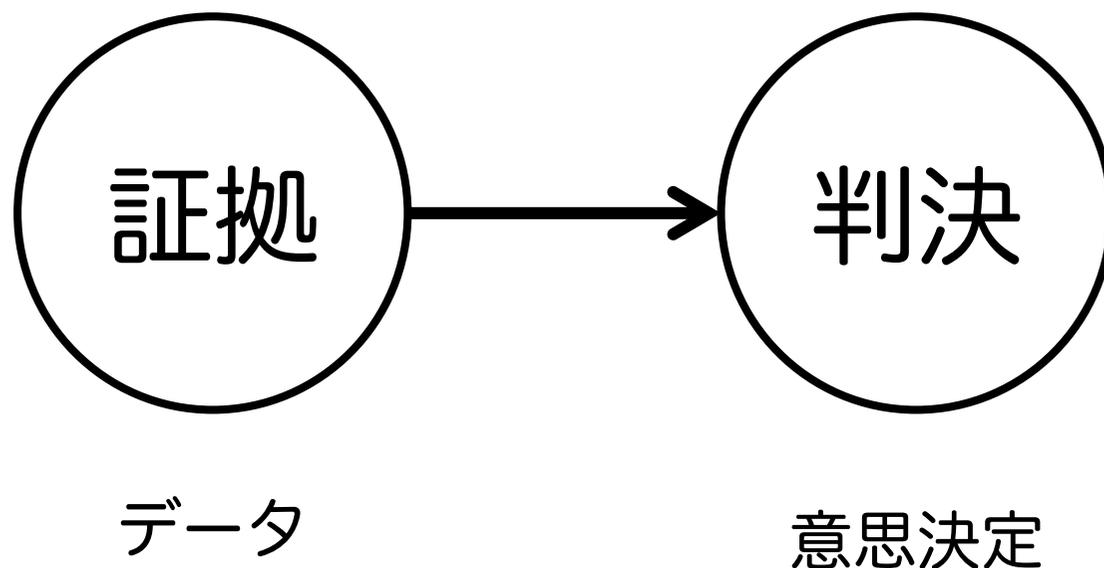
2014年4月27日(日)、5月4日(日)、11日(日) 午前0時30分～[*土曜深夜]

緊急ニュース等のため放送時間が変更になる場合があります。最新の放送予定はNHK番組表でご確認ください。

最新の放送予定
こちらをクリック!

<http://www9.nhk.or.jp/kaigai/sherlock2/>

裁判における意思決定



証拠があるという条件下で有罪にできる可能性は？

Baysian analysis in military operations

- 洋上で紛失した武器の捜索にベイズ理論が用いられた
- 紛失した確率をベイズ理論により推測
- 近年では無人機やロボットを活用したオペレーションにベイズ統計が応用

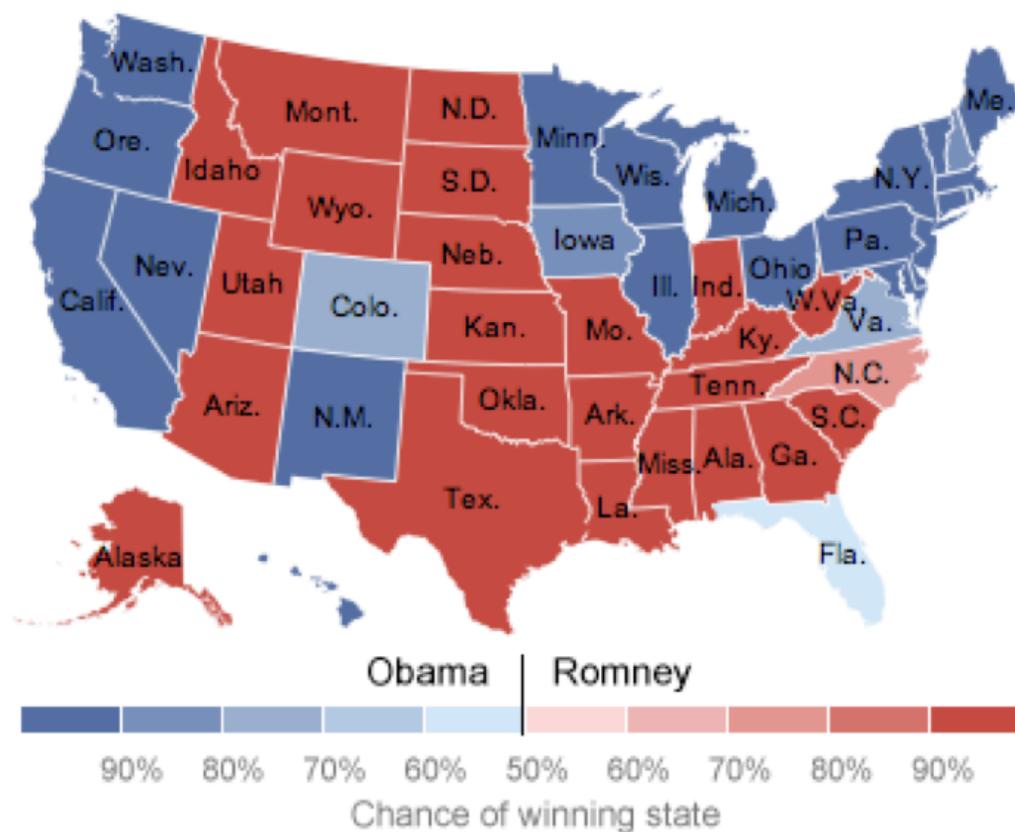


U.S. Army Research Laboratory
<http://www.arl.army.mil/www/default.cfm?page=185>

米大統領選(2012)と得票率

Nate Silver's Predictions on Election Eve

State-by-State Probabilities



<http://fivethirtyeight.blogs.nytimes.com/>



授業スケジュール

第01回	イントロダクション	第08回	ベイズ推定 (1)
第02回	確率 (1)	第09回	ベイズ推定 (2)
第03回	確率 (2)	第10回	MCMC (1)
第04回	ベイズの定理 (1)	第11回	MCMC (2)
第05回	ベイズの定理 (2)	第12回	ベイズ統計データ分析 (1)
第06回	ベイズの定理の応用 (1)	第13回	ベイズ統計データ分析 (2)
第07回	ベイズの定理の応用 (2)	第14回	ベイズ統計データ分析 (3)

ソフトウェア

- 授業の演習にはRというフリー統計ソフトを使います。以下のサイトからダウンロード可能ですので各自インストールしてください。

<http://cran.r-project.org/>



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled

成績判定方法

- 出欠点（各回の内容に関するクイズ）
- 期末レポート

主要参考文献

- 安道 知寛 (2010) 『ベイズ統計モデリング』、朝倉書店
- 照井 伸彦 (2010) 『Rによるベイズ統計分析』、朝倉書店
- 涌井良幸 (2009) 『道具としてのベイズ統計』、日本実業出版社
- 古谷知之 (2008) 『ベイズ統計データ分析』、朝倉書店
- 松原望 (2008) 『入門ベイズ統計』、東京図書
- Albert, J. (2007) "Bayesian Computation with R", Springer.
- Martin, J-M. and Robert, C.P. (2007) "Bayesian Core: A Practical Approach to Computational Bayesian Statistics", Springer.
- Koop, G., Poirier, D.J. and Tobias, J.L. (2007) "Bayesian Econometric Methods", Cambridge University Press.
- Lancaster, T. (2006) "An Introduction to Modern Bayesian Econometrics", Blackwell.
- Gelman, A. and Hill, J. (2006) "Data Analysis Using Regression and Multilevel/Hierarchical Models", Cambridge University Press.
- Congdon, P. (2006) "Bayesian Statistical Modelling" (2nd edition), Wiley.
- Congdon, P. (2003) "Applied Bayesian Modelling", Wiley.

データサイエンス・コンサルタント

- データサイエンス科目の学習を支援するTA(teaching assistant)とSA(student assistant)のグループ
- 1(イオ)202研究室で、データサイエンス科目履修や授業内容などについて学生の質問に答えてくれます
- Web: <http://ds.sfc.keio.ac.jp>
- 連絡先: dsc@sfc.keio.ac.jp